CIRJE-F-590

Bartlett-type Correction of the Generalized Least Squares Test in the Fay-Herriot Model

Tatsuya Kubokawa
University of Tokyo

September 2008; Revised in June 2009

CIRJE Discussion Papers can be downloaded without charge from:

http://www.e.u-tokyo.ac.jp/cirje/research/03research02dp.html

Discussion Papers are a series of manuscripts in their draft form. They are not intended for circulation or distribution except as indicated by the author. For that reason Discussion Papers may not be reproduced or distributed without the written consent of the author.

Bartlett-type Correction of the Generalized Least Squares Test in the Fay-Herriot Model

Tatsuya Kubokawa
University of Tokyo

June 26, 2009

Abstract

Consider the problem of testing the linear hypothesis on the regression coefficients in the Fay-Herriot model which has been used in the small area problem. Since this model involves the random effects, a test based on the generalized least squares estimator, called the GLS test, depends on the estimate of the 'between' component of variance, which causes the problem that it has an inflated type I error (size) when the variance component is far from zero. To fix this problem, we derive the second order approximation of the distribution of the GLS test statistic under the null hypothesis. Using the Bartlett-type correction, we obtain modified test statistics with sizes identical to the nominal significance level in the second-order asymptotic. As estimators of the variance component, the Prasad-Rao estimator, Fay-Herriot estimator, maximum likelihood (ML) estimator and the restricted maximum likelihood (REML) estimator are used and the corresponding modified tests based on the Bartlett-type correction are given. The sizes of these tests are investigated numerically and the Bartlett-type correction is shown to work well.

Key word and Phrases: Bartlett correction, Fay-Herriot model, generalized least square method, linear hypothesis, linear mixed model, maximum likelihood estimator, restricted maximum likelihood estimator, random effect, test, type I error, variance component.

1 Introduction

In the linear mixed models (LMM), we are concerned with the problem of testing linear hypotheses on the regression coefficients. It is known that the naive F-statistic based on the ordinary least squares (OLS) estimator has the serious drawback of having an inflated type I error (size) when the random effect is present. To fix the drawback, Wu, Holt and Holmes (1988) and Rao, Sutradhar and Yue (1993) suggested alternative test procedures. Although their procedures give significant improvements in terms of type I errors, the sizes are still slightly larger than the nominal significance level. In this paper, we derive the Bartlett-type correction in the Fay-Herriot model so that the resulting corrected tests have sizes which are identical to the nominal level upto the second order.

More specifically, we consider the model treated by Fay and Herriot (1979), given by

$$y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + v_i + e_i, \quad i = 1, \ldots, k, \tag{1.1}$$

where $\boldsymbol{\beta}$ is a $p \times 1$ vector of regression coefficients. It is assumed that $v_1, \ldots, v_k$ and $e_1, \ldots, e_k$ are mutually independently distributed as $v_i \sim \mathcal{N}(0, \psi)$ and $e_i \sim \mathcal{N}(0, d_i)$ for $i = 1, \ldots, k$. In the Fay-Herriot model, it is also assumed that d_i's are known constants. This model has been used as an area-level model in the context of the small area problem. Letting $\boldsymbol{y} = (y_1, \ldots, y_k)'$, $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k)'$, $\boldsymbol{v} = (v_1, \ldots, v_k)'$ and $\boldsymbol{e} = (e_1, \ldots, e_k)'$, we can write the model in the matrix notation as

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{v} + \boldsymbol{e}, \tag{1.2}$$

where $\boldsymbol{y} \sim \mathcal{N}_k(\boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}(\psi))$ for

$$\boldsymbol{\Sigma}(\psi) = \psi \boldsymbol{I}_k + \boldsymbol{D}, \quad \boldsymbol{D} = \mathrm{diag}\,(d_1, \ldots, d_k).$$

It is assumed that $k \geq p$ and $\boldsymbol{X}$ is of full rank.

In this paper, we consider the problem of testing the linear hypothesis on the regression coefficients $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_{p-1})'$, namely

$$H_0 : \boldsymbol{C}\boldsymbol{\beta} = \boldsymbol{b} \quad versus \quad H_1 : \boldsymbol{C}\boldsymbol{\beta} \neq \boldsymbol{b},$$

where $\boldsymbol{C}$ is a $q \times p$ known matrix with full rank ($p \geq q$) and $\boldsymbol{b}$ is a $q \times 1$ known vector. When ψ is known, the generalized least squares (GLS) estimator of $\boldsymbol{\beta}$ is

$$\widehat{\boldsymbol{\beta}}(\psi) = \left\{\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}(\psi)\boldsymbol{X}\right\}^{-1} \boldsymbol{X}'\boldsymbol{\Sigma}^{-1}(\psi)\boldsymbol{y}, \tag{1.3}$$

which has $\mathcal{N}_p(\boldsymbol{\beta}, \{\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}(\psi)\boldsymbol{X}\}^{-1})$, and an exact test statistic based on the GLS $\widehat{\boldsymbol{\beta}}(\psi)$ is given by

$$T_{GLS}(\psi) = \left\{\boldsymbol{C}\widehat{\boldsymbol{\beta}}(\psi) - \boldsymbol{b}\right\}' \left\{\boldsymbol{C}\left\{\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}(\psi)\boldsymbol{X}\right\}^{-1}\boldsymbol{C}'\right\}^{-1} \left\{\boldsymbol{C}\widehat{\boldsymbol{\beta}}(\psi) - \boldsymbol{b}\right\}.$$

which is distributed as χ^2_q distribution with q degrees of freedom under H_0. Thus, the rejection region with size $0 < \alpha < 1$ is expressed as $\{\boldsymbol{y} | T_{GLS}(\psi) > \chi^2_{q,\alpha}\}$, where $\chi^2_{q,\alpha}$ is the upper $100\alpha\%$ point of the χ^2_q distribution.

When the model (1.1) does not include the random effects v_i's, $T_{GLS}(0)$ is an exact test, namely, $P[T_{GLS}(0) > \chi^2_{q,\alpha}] = \alpha$ under H_0. Since the random effects are present in the model (1.1), however, $T_{GLS}(0)$ has the serious drawback of having an inflated type I error when ψ is away from zero. An alternative procedure is the test statistic $T_{GLS}(\widehat{\psi})$ where $\widehat{\psi}$ is an appropriate estimator of ψ, and we call it the Generalized Least Squares (GLS) test. Although the GLS test $T_{GLS}(\widehat{\psi})$ is a reasonable alternative, as shown in Section 3, the size of $T_{GLS}(\widehat{\psi})$ remains still slightly larger than the nominal significance levels.

To improve on $T_{GLS}(\widehat{\psi})$ in the sense of controlling the size, we derive the Bartlett-type correction of $T_{GLS}(\widehat{\psi})$ when k is large, and show that the resulting corrected test $T^*_{GLS}(\widehat{\psi})$

has the size which is identical to the nominal significance level α up to $O(k^{-1})$ under the null hypothesis H_0, namely, $P[T^*_{GLS}(\widehat{\psi}) > \chi^2_{q,\alpha}] = \alpha + o(k^{-1})$ as $k \to \infty$ under H_0.

In Section 2, we obtain the asymptotic expansion of the distribution of $T_{GLS}(\widehat{\psi})$ under the null hypothesis H_0 and provide the Bartlett-type correction of $T_{GLS}(\widehat{\psi})$ using the monotone transformation given in Fujikoshi (2000). The resulting corrected test can be shown to satisfy the size property in the second order asymptotic on k. The Bartlett-type correction depends on the asymptotic bias and variance of the estimator $\widehat{\psi}$ of ψ. Section 2 also introduces some estimators of ψ and provides their asymptotic bias and variance as well as it is shown that the estimators satisfy the assumptions required to establish the asymptotic expansion of $T_{GLS}(\widehat{\psi})$. In Section 3, the sizes of the corrected tests are numerically investigated for various k and ψ, and it is shown that the corrected tests improve on $T_{GLS}(\widehat{\psi})$ in the sense of controlling the sizes. The proofs of the main results are given in Section 4.

2 Main results

2.1 Bartlett-type correction of the GLS test

In the Fay-Herriot model (1.1) or (1.2), we consider the problem of testing the linear hypothesis on the regression coefficients $\boldsymbol{\beta}$, given by $H_0 : \boldsymbol{C\beta} = \boldsymbol{b}$ against $H_1 : \boldsymbol{C\beta} \neq \boldsymbol{b}$, where $\boldsymbol{C}$ is a $q \times p$ known matrix with full rank $(p \geq q)$ and $\boldsymbol{b}$ is a $q \times 1$ known vector. Since $\boldsymbol{y} \sim \mathcal{N}_k(\boldsymbol{X\beta}, \boldsymbol{\Sigma}(\psi))$ for $\boldsymbol{\Sigma}(\psi) = \psi \boldsymbol{I}_k + \boldsymbol{D}$, the GLS estimator of $\boldsymbol{\beta}$ is $\widehat{\boldsymbol{\beta}}(\psi) = \{\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}(\psi)\boldsymbol{X}\}^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}(\psi)\boldsymbol{y}$. When ψ is known, the chi-square test statistic based on $\widehat{\boldsymbol{\beta}}(\psi)$ is given by

$$T_{GLS}(\psi) = \left\{\boldsymbol{C}\widehat{\boldsymbol{\beta}}(\psi) - \boldsymbol{b}\right\}' \left\{\boldsymbol{C}\left\{\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}(\psi)\boldsymbol{X}\right\}^{-1}\boldsymbol{C}'\right\}^{-1} \left\{\boldsymbol{C}\widehat{\boldsymbol{\beta}}(\psi) - \boldsymbol{b}\right\}. \tag{2.1}$$

which is distributed as χ^2_q distribution with q degrees of freedom under H_0. Thus, the critical region is expressed as $\{\boldsymbol{y} | T_{GLS}(\psi) > \chi^2_{q,\alpha}\}$, where $\chi^2_{q,\alpha}$ is the upper $100\alpha\%$ point of the χ^2_q distribution. Since ψ is unknown, however, we need to estimate ψ by an appropriate estimator $\widehat{\psi}$ and consider the test statistic $T_{GLS}(\widehat{\psi})$ by substituting $\widehat{\psi}$ into $T_{GLS}(\psi)$. Since $T_{GLS}(\widehat{\psi})$ is based on the GLS estimator $\widehat{\boldsymbol{\beta}}(\psi)$, we call it the GLS test. This substitution causes the problem that the size of the GLS test $T_{GLS}(\widehat{\psi})$ is not identical to the nominal level α. As seen in Tables 1-3 in Section 3, the problem is serious when k is small and ψ is away from zero. To improve the GLS test in the sense of controlling the size, we derive the Bartlett-type correction of the GLS test. To this end, we first obtain the asymptotic expansion of the GLS test under the null hypothesis H_0.

Let $\boldsymbol{W}(\psi) = \boldsymbol{C}'\{\boldsymbol{C}\left\{\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}(\psi)\boldsymbol{X}\right\}^{-1}\boldsymbol{C}'\}^{-1}\boldsymbol{C}$ and $\boldsymbol{W}^{(i)}(\psi) = (\mathrm{d}^i/\mathrm{d}\psi^i)\boldsymbol{W}(\psi)$ for $i = 1, 2$. Also let $\boldsymbol{A}_i(\psi) = \boldsymbol{X}'\boldsymbol{\Sigma}^{-i}(\psi)\boldsymbol{X}$ for $i = 1, 2, 3$. As the differential operator with respect to $\boldsymbol{y}$, we use the notation

$$\boldsymbol{\nabla} = \frac{\partial}{\partial \boldsymbol{y}} = \left(\frac{\partial}{\partial y_1}, \ldots, \frac{\partial}{\partial y_k}\right)'.$$

Assume the following conditions:

($A1$) $\boldsymbol{X}$ is of full rank, $k \geq p$ and $\boldsymbol{X}'\boldsymbol{X} = O(k)$ as $k \to \infty$.

($A2$) The estimator $\widehat{\psi}$ satisfies that $\boldsymbol{X}'\boldsymbol{\nabla}\widehat{\psi} = \boldsymbol{0}$ and $\boldsymbol{X}'\boldsymbol{\nabla}\boldsymbol{\nabla}'\widehat{\psi} = \boldsymbol{0}$.

($A3$) $\widehat{\psi} - \psi = O_p(k^{-1/2})$, $\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}(\psi)\boldsymbol{\nabla}\widehat{\psi} = O_p(k^{-1/2})$ and $\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}(\psi)\{\boldsymbol{\nabla}\boldsymbol{\nabla}'\widehat{\psi}\}\boldsymbol{\Sigma}^{-1}(\psi)\boldsymbol{X} = O_p(1)$ as $k \to \infty$.

($A4$) $Bias(\widehat{\psi}) = O(k^{-1})$ as $k \to \infty$.

($A5$) $\widehat{\psi}$ is invariant under the transformation $\boldsymbol{y} \to \boldsymbol{y} + \boldsymbol{X}\boldsymbol{\alpha}$ for any $\boldsymbol{\alpha} \in \boldsymbol{R}^p$.

Theorem 2.1 *Assume the conditions* $(A1) - (A5)$*, Under the null hypothesis* H_0*, the distribution function of the GLS test* $T_{GLS}(\widehat{\psi})$ *is approximated as*

$$
\begin{aligned}
P[T_{GLS}(\widehat{\psi}) \leq x] =& G_q(x) + \{G_{q+2}(x) - G_q(x)\}\, h_1(\psi) \\
&+ \{G_{q+4}(x) - 2G_{q+2} + G_q(x)\}\, h_2(\psi) + o(k^{-1}) \qquad (2.2) \\
=& G_q(x) - 2\frac{x}{q}\left\{h_1(\psi) - h_2(\psi) + \frac{x}{q+2}h_2(\psi)\right\} g_q(x) + o(k^{-1}), \qquad (2.3)
\end{aligned}
$$

where $G_q(x)$ *and* $g_q(x)$ *are the distribution and density functions of* χ^2_q*, the chi-square distribution with* q *degrees of freedom, and*

$$
\begin{aligned}
h_1(\psi) =& \frac{1}{2}\mathrm{tr}\,[\boldsymbol{W}^{(1)}(\psi)\boldsymbol{A}_1^{-1}(\psi)]Bias(\widehat{\psi}) + \frac{1}{2}\Big\{\frac{1}{2}\mathrm{tr}\,[\boldsymbol{W}^{(2)}(\psi)\boldsymbol{A}_1^{-1}(\psi)] \\
&+ \mathrm{tr}\,[\boldsymbol{A}_1^{-1}(\psi)\boldsymbol{W}(\psi)\boldsymbol{A}_1^{-1}(\psi)\{\boldsymbol{A}_3(\psi) - \boldsymbol{A}_2(\psi)\boldsymbol{A}_1^{-1}(\psi)\boldsymbol{A}_2(\psi)\}]\Big\}Var(\widehat{\psi}), \qquad (2.4) \\
h_2(\psi) =& \frac{1}{8}\Big\{\Big(\mathrm{tr}\,[\boldsymbol{W}^{(1)}(\psi)\boldsymbol{A}_1^{-1}(\psi)]\Big)^2 + 2\mathrm{tr}\,\Big[\boldsymbol{W}^{(1)}(\psi)\boldsymbol{A}_1^{-1}(\psi)\Big]^2\Big\}Var(\widehat{\psi}).
\end{aligned}
$$

From (2.2), it can be seen that the first and second moments of $T_{GLS}(\widehat{\psi})$ are given by

$$
\begin{aligned}
E[T_{GLS}(\widehat{\psi})] =& q\left\{1 + k^{-1}c_1(\psi)\right\} + o(k^{-1}), \\
E[\{T_{GLS}(\widehat{\psi})\}^2] =& q(q+2)\left\{1 + k^{-1}c_2(\psi)\right\} + o(k^{-1}),
\end{aligned}
$$

where

$$
\begin{aligned}
c_1(\psi) =& 2kh_1(\psi)/q, \\
c_2(\psi) =& 4k\left\{(q+2)h_1(\psi) + 2h_2(\psi)\right\}/\{q(q+2)\}.
\end{aligned}
$$

Then, we can use the results of Fujikoshi (2000) to derive the Bartlett-type correction of $T_{GLS}(\widehat{\psi})$. Let $\tilde{c}_2(\psi) = c_2(\psi) - 2c_1(\psi)$,

$$
\begin{aligned}
\alpha(\psi) =& 2/\tilde{c}_2(\psi), \\
\beta(\psi) =& \frac{1}{2}\left\{(q+2)c_2(\psi) - 2(q+4)c_1(\psi)\right\}/\tilde{c}_2(\psi) \\
=& \frac{4k}{q\tilde{c}_2(\psi)}\left\{h_2(\psi) - h_1(\psi)\right\}.
\end{aligned}
$$

Then, it is observed that $\tilde{c}_2(\psi) = 8kh_2(\psi)/\{q(q+2)\} > 0$, namely, $\alpha(\psi) > 0$, and

$$k\alpha(\psi) + \beta(\psi) = \frac{2k}{q\tilde{c}_2(\psi)} \left[q + 2\left\{h_2(\psi) - h_1(\psi)\right\}\right].$$

Fujikoshi (2000) considered the following monotone transformations:

$$T_1^*(\widehat{\psi}) = \left\{k\alpha(\widehat{\psi}) + \beta(\widehat{\psi})\right\} \log\left(1 + \frac{1}{k\alpha(\widehat{\psi})} T_{GLS}(\widehat{\psi})\right), \tag{2.5}$$

$$T_2^*(\widehat{\psi}) = \left\{k\alpha(\widehat{\psi}) + \beta(\widehat{\psi})\right\} \left\{1 - \exp\left(-\frac{1}{k\alpha(\widehat{\psi})} T_{GLS}(\widehat{\psi})\right)\right\}, \tag{2.6}$$

where $T_1^*(\widehat{\psi})$ is defined when $k\alpha(\widehat{\psi}) + \beta(\widehat{\psi}) > 0$, while $T_2^*(\widehat{\psi})$ can be defined whenever $k\alpha(\widehat{\psi}) + \beta(\widehat{\psi})$ takes any value. It is noted that both transformations have the same expansion up to $O(k^{-1})$, given by

$$T_i^*(\widehat{\psi}) = T_{GLS}(\widehat{\psi}) + \frac{1}{k\alpha(\widehat{\psi})} \left\{\beta(\widehat{\psi})T_{GLS}(\widehat{\psi}) - 2^{-1}(T_{GLS}(\widehat{\psi}))^2\right\} + O(k^{-2}),$$

for $i = 1, 2$, which implies that

$$E[T_i^*(\widehat{\psi})] = q + o(k^{-1}) \quad \text{and} \quad E[\{T_i^*(\widehat{\psi})\}^2] = q(q+2) + o(k^{-1}).$$

From Theorem 2 of Fujikoshi (2000), we can see that the sizes of the modified test statistics with the Bartlett-type correction are justified theoretically.

Theorem 2.2 (Bartlett-type Correction) *Assume the conditions* (A1)$-$(A5). *Then, the sizes of the corrected tests* $T_i^*(\widehat{\psi})$ *given by* (2.5) *and* (2.6) *are identical to the nominal significance level* α *up to* $O(k^{-1})$ *as* $k \to \infty$, *namely,*

$$P[T_i^*(\widehat{\psi}) > \chi^2_{q,\alpha}] = \alpha + o(k^{-1}),$$

under H_0.

For another account on monotone transformations, see Fujisawa (1997), Enoki and Aoshima (2006) and the references therein. It is noted that the conventional Bartlett correction based on only the first moment $E[T_{GLS}(\widehat{\psi})]$ is given by

$$T_B(\widehat{\psi}) = T_{GLS}(\widehat{\psi})/(1 + 2q^{-1}h_1(\widehat{\psi})), \tag{2.7}$$

but the size of $T_B(\widehat{\psi})$ does not possess the same asymptotic approximation as in Theorem 2.2 unless $c_2(\psi) = 0$.

When $\boldsymbol{C} = \boldsymbol{I}$, the functions $h_1(\psi)$ and $h_2(\psi)$ can be simplified. In this case, $q = p$, $\boldsymbol{W}(\psi) = \boldsymbol{A}_1(\psi)$, $\boldsymbol{W}^{(1)}(\psi) = \boldsymbol{A}_1^{(1)}(\psi) = -\boldsymbol{A}_2(\psi)$ and $\boldsymbol{W}^{(2)}(\psi) = \boldsymbol{A}_1^{(2)}(\psi) = 2\boldsymbol{A}_3(\psi)$, so that $h_1(\psi)$ and $h_2(\psi)$ are expressed as

$$\begin{aligned} h_1(\psi) = &- \frac{1}{2}\mathrm{tr}\left[\boldsymbol{A}_1^{-1}(\psi)\boldsymbol{A}_2(\psi)\right] Bias_\psi(\widehat{\psi}) \\ &+ \frac{1}{2}\left\{2\mathrm{tr}\left[\boldsymbol{A}_1^{-1}(\psi)\boldsymbol{A}_3(\psi)\right] - \mathrm{tr}\left[\boldsymbol{A}_1^{-1}(\psi)\boldsymbol{A}_2(\psi)\right]^2\right\} Var_\psi(\widehat{\psi}), \end{aligned}$$

and

$$h_2(\psi) = \frac{1}{8}\left\{\left(\mathrm{tr}\,[\boldsymbol{A}_2(\psi)\boldsymbol{A}_1^{-1}(\psi)]\right)^2 + 2\mathrm{tr}\,\left[\boldsymbol{A}_2(\psi)\boldsymbol{A}_1^{-1}(\psi)\right]^2\right\} Var(\widehat{\psi}).$$

For the general matrix $\boldsymbol{C}$, $\boldsymbol{W}^{(2)}(\psi)$ gives a complicated form, but we can compute it as follows: Let $\boldsymbol{E}(\psi) = \{\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}(\psi)\boldsymbol{X}\}^{-1}$ and $\boldsymbol{F}(\psi) = \{\boldsymbol{C}\boldsymbol{E}(\psi)\boldsymbol{C}'\}^{-1}$. Then, $\boldsymbol{W}^{(1)}(\psi) = \boldsymbol{C}'\boldsymbol{F}^{(1)}(\psi)\boldsymbol{C}$, which can be obtained from $\boldsymbol{F}^{(1)}(\psi) = -\boldsymbol{F}(\psi)\boldsymbol{C}\boldsymbol{E}^{(1)}(\psi)\boldsymbol{C}'\boldsymbol{F}(\psi)$ and $\boldsymbol{E}^{(1)}(\psi) = \boldsymbol{E}(\psi)\boldsymbol{A}_2\boldsymbol{E}(\psi)$. Also, we have $\boldsymbol{W}^{(2)}(\psi) = \boldsymbol{C}'\boldsymbol{F}^{(2)}(\psi)\boldsymbol{C}$, which can be obtained from

$$\boldsymbol{F}^{(2)}(\psi) = -\boldsymbol{F}^{(1)}(\psi)\boldsymbol{C}\boldsymbol{E}^{(1)}(\psi)\boldsymbol{C}'\boldsymbol{F}(\psi) - \boldsymbol{F}(\psi)\boldsymbol{C}\boldsymbol{E}^{(2)}(\psi)\boldsymbol{C}'\boldsymbol{F}(\psi) - \boldsymbol{F}(\psi)\boldsymbol{C}\boldsymbol{E}^{(1)}(\psi)\boldsymbol{C}'\boldsymbol{F}^{(1)}(\psi)$$

and

$$\boldsymbol{E}^{(2)}(\psi) = -2\boldsymbol{E}(\psi)\boldsymbol{A}_2(\psi)\boldsymbol{E}(\psi)\boldsymbol{A}_2(\psi)\boldsymbol{E}(\psi) + 2\boldsymbol{E}(\psi)\boldsymbol{A}_3(\psi)\boldsymbol{E}(\psi).$$

2.2 Estimators of ψ and the corresponding corrections of the GLS test

The Bartlett-type corrections $T_i^*(\widehat{\psi})$ given in (2.5) and (2.6) depend on the asymptotic bias and variance of the estimator $\widehat{\psi}$. We here provide these values for some specific estimators of ψ as well as we check the assumptions $(A2)$-$(A5)$ on $\widehat{\psi}$. The estimators we treat are the estimator proposed by Prasad and Rao (1990), the estimator proposed by Fay and Herriot (1979), the maximum likelihood estimator and the restricted maximum likelihood estimator.

[1] Prasad-Rao estimator. Prasad and Rao (1990) proposed the estimator

$$\widehat{\psi}_{PR} = \frac{1}{k-p}(\boldsymbol{y}'\boldsymbol{Q}_0\boldsymbol{y} - \mathrm{tr}\,[\boldsymbol{D}\boldsymbol{Q}_0])_+, \tag{2.8}$$

where $(a)_+ = \max\{a, 0\}$ and

$$\boldsymbol{Q}_0 = \boldsymbol{I}_k - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'.$$

Since $\boldsymbol{Q}_0\boldsymbol{X} = \boldsymbol{0}$, the condition $(A5)$ is satisfied. Since $P[\widehat{\psi}_{PR} = 0] = o(k^{-1})$, it is sufficient to handle the case that $\widehat{\psi}_{PR} > 0$. It is noted that $(\boldsymbol{y}'\boldsymbol{Q}_0\boldsymbol{y} - \mathrm{tr}\,[\boldsymbol{D}\boldsymbol{Q}_0])/(k-p)$ is an unbiased estimator of ψ. Note that

$$\boldsymbol{\nabla}\widehat{\psi}_{PR} = 2(k-p)^{-1}\boldsymbol{Q}_0\boldsymbol{y}.$$

Then it is observed that $\boldsymbol{X}'\boldsymbol{\nabla}\widehat{\psi}_{PR} = \boldsymbol{0}$ and $\boldsymbol{\nabla}\boldsymbol{\nabla}'\widehat{\psi}_{PR} = 2(k-p)^{-1}\boldsymbol{Q}_0$, so that for $\boldsymbol{A}_i(\psi) = \boldsymbol{X}'\boldsymbol{\Sigma}^{-i}(\psi)\boldsymbol{X}$, $i = 1, 2$,

$$\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}(\psi)\boldsymbol{\nabla}\widehat{\psi}_{PR} = \frac{2}{k-p}\boldsymbol{A}_1(\psi)\{(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}(\psi)\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}(\psi)\boldsymbol{y} - (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}\},$$

which is $O_p(k^{-1/2})$, and

$$\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}(\psi)\{\boldsymbol{\nabla}\boldsymbol{\nabla}'\widehat{\psi}_{PR}\}\boldsymbol{\Sigma}^{-1}(\psi)\boldsymbol{X} = \frac{2}{k-p}\{\boldsymbol{A}_2(\psi) - \boldsymbol{A}_1(\psi)(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}_1(\psi)\},$$

which is $O(1)$. Prasad and Rao (1990) demonstrated that

$$\begin{aligned} Var(\widehat{\psi}_{PR}) =& 2k^{-2}\text{tr}\,\boldsymbol{\Sigma}^2(\psi) + o(k^{-1}), \\ Bias(\widehat{\psi}_{PR}) =& o(k^{-1}), \end{aligned}$$

so that the assumptions $(A2)$-$(A4)$ are satisfied, and from Theorem 2.2, we get the Bartlett-type correction.

[2] Fay-Herriot estimator. Fay and Herriot (1978) proposed the estimator of the form $\widehat{\psi}_{FH} = (\widehat{\psi}^*_{FH})_+$, where $\widehat{\psi}^*_{FH}$ is a solution of the equation

$$\boldsymbol{y}'\boldsymbol{Q}_1(\widehat{\psi}^*_{FH})\boldsymbol{y} = k - p, \tag{2.9}$$

for

$$\boldsymbol{Q}_1(\psi) = \left\{\boldsymbol{I}_k - \boldsymbol{\Sigma}^{-1}(\psi)\boldsymbol{X}\boldsymbol{A}_1^{-1}(\psi)\boldsymbol{X}'\right\}\boldsymbol{\Sigma}^{-1}(\psi)\left\{\boldsymbol{I}_k - \boldsymbol{X}\boldsymbol{A}_1^{-1}(\psi)\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}(\psi)\right\}.$$

Since $\boldsymbol{Q}_1(\psi)\boldsymbol{X} = \boldsymbol{0}$, the condition $(A5)$ is satisfied. Since $P[\widehat{\psi}_{FH} = 0] = o(k^{-1})$, it is sufficient to handle the case that $\widehat{\psi}_{FH} > 0$. Note that

$$\boldsymbol{\nabla}\widehat{\psi}_{FH} = -\frac{2\boldsymbol{Q}_1(\widehat{\psi}_{FH})\boldsymbol{y}}{\boldsymbol{y}'\boldsymbol{Q}_1^{(1)}(\widehat{\psi}_{FH})\boldsymbol{y}}.$$

Then it can be seen that $\boldsymbol{X}'\boldsymbol{\nabla}\widehat{\psi}_{FH} = \boldsymbol{0}$, $\boldsymbol{y}'\boldsymbol{Q}_1^{(1)}(\widehat{\psi}_{FH})\boldsymbol{y} = O_p(k)$ and $\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}(\psi)\boldsymbol{\nabla}\widehat{\psi}_{FH} = O_p(k^{-1/2})$. Denote $\boldsymbol{Q}_1^{(2)}(\psi) = (\mathrm{d}^2/\mathrm{d}\psi^2)\boldsymbol{Q}_1(\psi)$. For $\widehat{\psi} = \widehat{\psi}_{FH}$, it is seen that

$$\boldsymbol{\nabla}\boldsymbol{\nabla}'\widehat{\psi}_{FH} = -\frac{2\boldsymbol{Q}_1(\widehat{\psi}) + 2\boldsymbol{Q}_1^{(1)}(\widehat{\psi})\boldsymbol{y}(\boldsymbol{\nabla}\widehat{\psi})' + (\boldsymbol{y}'\boldsymbol{Q}_1^{(2)}(\widehat{\psi})\boldsymbol{y})(\boldsymbol{\nabla}\widehat{\psi})(\boldsymbol{\nabla}\widehat{\psi})'}{\boldsymbol{y}'\boldsymbol{Q}_1^{(1)}(\widehat{\psi})\boldsymbol{y}}.$$

It can be observed that $\boldsymbol{X}'\boldsymbol{\nabla}\boldsymbol{\nabla}'\widehat{\psi}_{FH} = \boldsymbol{0}$ and $\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}(\psi)\{\boldsymbol{\nabla}\boldsymbol{\nabla}'\widehat{\psi}_{FH}\}\boldsymbol{\Sigma}^{-1}(\psi)\boldsymbol{X} = O_p(1)$. Datta, Rao and Smith (2005) showed that

$$\begin{aligned} Var(\widehat{\psi}_{FH}) =& 2k\left\{\text{tr}\,\boldsymbol{\Sigma}^{-1}(\psi)\right\}^{-2} + o(k^{-1}), \\ Bias(\widehat{\psi}_{FH}) =& 2\frac{k\text{tr}\,\boldsymbol{\Sigma}^{-2}(\psi) - \{\text{tr}\,\boldsymbol{\Sigma}^{-1}(\psi)\}^2}{\{\text{tr}\,\boldsymbol{\Sigma}^{-1}(\psi)\}^3} + o(k^{-1}), \end{aligned}$$

so that the assumptions $(A2)$-$(A4)$ are satisfied and we get the Bartlett-type correction from Theorem 2.2.

[3] Maximum likelihood (ML) estimator. The ML estimator is given by $\widehat{\psi}_{ML} = (\widehat{\psi}^*_{ML})_+$ where $\widehat{\psi}^*_{ML}$ is a solution of the equation

$$\boldsymbol{y}'\boldsymbol{Q}_2(\widehat{\psi}^*_{ML})\boldsymbol{y} = \text{tr}\,[\boldsymbol{\Sigma}^{-1}(\widehat{\psi}^*_{ML})], \tag{2.10}$$

for

$$\boldsymbol{Q}_2(\psi) = \left\{\boldsymbol{I}_k - \boldsymbol{\Sigma}^{-1}(\psi)\boldsymbol{X}\boldsymbol{A}_1^{-1}(\psi)\boldsymbol{X}'\right\}\boldsymbol{\Sigma}^{-2}(\psi)\left\{\boldsymbol{I}_k - \boldsymbol{X}\boldsymbol{A}_1^{-1}(\psi)\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}(\psi)\right\}.$$

Since $\boldsymbol{Q}_2(\psi)\boldsymbol{X} = \boldsymbol{0}$, the condition $(A5)$ is satisfied. Since $P[\widehat{\psi}_{ML} = 0] = o(k^{-1})$, it is sufficient to handle the case that $\widehat{\psi}_{ML} > 0$. Note that

$$\boldsymbol{\nabla}\widehat{\psi}_{ML} = -2\frac{\boldsymbol{Q}_2(\widehat{\psi}_{ML})\boldsymbol{y}}{\boldsymbol{y}'\boldsymbol{Q}_2^{(1)}(\widehat{\psi}_{ML})\boldsymbol{y} + \mathrm{tr}\,[\boldsymbol{\Sigma}^{-2}(\widehat{\psi}_{ML})]}.$$

Then for $\widehat{\psi} = \widehat{\psi}_{ML}$, it can be seen that $\boldsymbol{X}'\boldsymbol{\nabla}\widehat{\psi} = \boldsymbol{0}$, $\boldsymbol{y}'\boldsymbol{Q}_2^{(1)}(\widehat{\psi})\boldsymbol{y} = O_p(k)$, $\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}(\psi)\boldsymbol{\nabla}\widehat{\psi} = O_p(k^{-1/2})$ and

$$\begin{aligned}
&\boldsymbol{\nabla}\boldsymbol{\nabla}'\widehat{\psi}_{ML}\\
&= -\frac{2\boldsymbol{Q}_2(\widehat{\psi}) + 2\boldsymbol{Q}_2^{(1)}(\widehat{\psi})\boldsymbol{y}(\boldsymbol{\nabla}\widehat{\psi})' - 2\mathrm{tr}\,[\boldsymbol{\Sigma}^{-3}](\boldsymbol{\nabla}\widehat{\psi})(\boldsymbol{\nabla}\widehat{\psi})' + (\boldsymbol{y}'\boldsymbol{Q}_2^{(2)}(\widehat{\psi})\boldsymbol{y})(\boldsymbol{\nabla}\widehat{\psi})(\boldsymbol{\nabla}\widehat{\psi})'}{\boldsymbol{y}'\boldsymbol{Q}_2^{(1)}(\widehat{\psi})\boldsymbol{y} + \mathrm{tr}\,[\boldsymbol{\Sigma}^{-2}(\widehat{\psi})]},
\end{aligned}$$

It can be observed that $\boldsymbol{X}'\boldsymbol{\nabla}\boldsymbol{\nabla}'\widehat{\psi}_{ML} = \boldsymbol{0}$ and $\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}(\psi)\{\boldsymbol{\nabla}\boldsymbol{\nabla}'\widehat{\psi}_{ML}\}\boldsymbol{\Sigma}^{-1}(\psi)\boldsymbol{X} = O_p(1)$. Datta and Lahiri (2000) derived that

$$\begin{aligned}
Var(\widehat{\psi}_{ML}) =& 2\left\{\mathrm{tr}\,\boldsymbol{\Sigma}^{-2}(\psi)\right\}^{-1} + o(k^{-1}),\\
Bias(\widehat{\psi}_{ML}) =& -\frac{\mathrm{tr}\,[\boldsymbol{A}_1^{-1}(\psi)\boldsymbol{A}_2(\psi)]}{\mathrm{tr}\,\boldsymbol{\Sigma}^{-2}(\psi)} + o(k^{-1}),
\end{aligned}$$

so that the assumptions $(A2)$-$(A4)$ are satisfied. Hence we get the Bartlett-type correction from Theorem 2.2.

[4] **Restricted maximum likelihood (REML) estimator.** The REML estimator is given by $\widehat{\psi}_{REML} = (\widehat{\psi}^*_{REML})_+$ where $\widehat{\psi}^*_{REML}$ is a solution of the equation

$$\boldsymbol{y}'\boldsymbol{Q}_2(\widehat{\psi}^*_{REML})\boldsymbol{y} = \mathrm{tr}\,[\boldsymbol{\Sigma}^{-1}(\widehat{\psi}^*_{REML})] - \mathrm{tr}\,[\boldsymbol{A}_1^{-1}(\widehat{\psi}^*_{REML})\boldsymbol{A}_2(\widehat{\psi}^*_{REML})], \tag{2.11}$$

for $\boldsymbol{Q}_2(\psi)$ defined above. Since $P[\widehat{\psi}_{REML} = 0] = o(k^{-1})$, it is sufficient to handle the case that $\widehat{\psi}_{REML} > 0$. Note that for $\widehat{\psi} = \widehat{\psi}_{REML}$,

$$\boldsymbol{\nabla}\widehat{\psi}_{REML} = -2\frac{\boldsymbol{Q}_2(\widehat{\psi})\boldsymbol{y}}{\boldsymbol{y}'\boldsymbol{Q}_2^{(1)}(\widehat{\psi})\boldsymbol{y} + \mathrm{tr}\,[\boldsymbol{\Sigma}^{-2}(\widehat{\psi})] + \mathrm{tr}\,[\boldsymbol{A}_1^{-1}(\widehat{\psi})\boldsymbol{A}_2(\widehat{\psi})]^2 - 2\mathrm{tr}\,\boldsymbol{A}_3(\widehat{\psi})}.$$

Then it can be seen that $\boldsymbol{X}'\boldsymbol{\nabla}\widehat{\psi}_{REML} = \boldsymbol{0}$ and $\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}(\psi)\boldsymbol{\nabla}\widehat{\psi}_{REML} = O_p(k^{-1/2})$. Also it is observed that for $\widehat{\psi} = \widehat{\psi}_{REML}$,

$$\begin{aligned}
&2\boldsymbol{Q}_2(\widehat{\psi}) + 2\boldsymbol{Q}_2^{(1)}(\widehat{\psi})\boldsymbol{y}(\boldsymbol{\nabla}\widehat{\psi})' - 2\mathrm{tr}\,[\boldsymbol{\Sigma}^{-3}(\widehat{\psi})](\boldsymbol{\nabla}\widehat{\psi})(\boldsymbol{\nabla}\widehat{\psi})' + 6\mathrm{tr}\,[\boldsymbol{A}_4(\widehat{\psi})](\boldsymbol{\nabla}\widehat{\psi})(\boldsymbol{\nabla}\widehat{\psi})'\\
&\quad + 2\left\{\mathrm{tr}\,[\boldsymbol{A}_1^{-1}(\widehat{\psi})\boldsymbol{A}_2(\widehat{\psi})]^3 - 2\mathrm{tr}\,[\boldsymbol{A}_1^{-1}(\widehat{\psi})\boldsymbol{A}_2(\widehat{\psi})\boldsymbol{A}_1^{-1}(\widehat{\psi})\boldsymbol{A}_3(\widehat{\psi})]\right\}(\boldsymbol{\nabla}\widehat{\psi})(\boldsymbol{\nabla}\widehat{\psi})'\\
&= -\left\{\boldsymbol{y}'\boldsymbol{Q}_2^{(1)}(\widehat{\psi})\boldsymbol{y} + \mathrm{tr}\,[\boldsymbol{\Sigma}^{-2}(\widehat{\psi})] + \mathrm{tr}\,[\boldsymbol{A}_1^{-1}(\widehat{\psi})\boldsymbol{A}_2(\widehat{\psi})]^2 - 2\mathrm{tr}\,\boldsymbol{A}_3(\widehat{\psi})\right\}\boldsymbol{\nabla}\boldsymbol{\nabla}'\widehat{\psi}.
\end{aligned}$$

It can be observed that $\boldsymbol{X}'\boldsymbol{\nabla}\boldsymbol{\nabla}'\widehat{\psi}_{REML} = \boldsymbol{0}$ and $\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}(\psi)\{\boldsymbol{\nabla}\boldsymbol{\nabla}'\widehat{\psi}_{REML}\}\boldsymbol{\Sigma}^{-1}(\psi)\boldsymbol{X} = O_p(1)$. Datta and Lahiri (2000) showed that $Var(\widehat{\psi}_{REML}) = Var(\widehat{\psi}_{ML}) + o(k^{-1})$ and $Bias(\widehat{\psi}_{REML}) = o(k^{-1})$, so that the assumptions $(A2)$-$(A4)$ are satisfied, and we get the Bartlett-type correction from Theorem 2.2.

3 Numerical Study

In this section, we shall investigate the performances of the sizes of the test statistics proposed in the previous section through simulation experiments.

For the regressor variables in the model (1.1), it is supposed that the model has an intercept term, namely, $\boldsymbol{x}_i' = (1, \boldsymbol{x}_i^{*\prime})$ for a $(p-1)$-vector $\boldsymbol{x}_i^*$, where $\boldsymbol{x}_i^*$ is generated as

$$\boldsymbol{x}_i^* = \boldsymbol{u} + \boldsymbol{z}_i.$$

Here, $\boldsymbol{z}_i$ is a $(p-1)$-random vector having $\mathcal{N}_{p-1}(\boldsymbol{0}, 10\boldsymbol{I}_{p-1})$, and $\boldsymbol{u}$ is a $(p-1)$-random vector having $\mathcal{N}_{p-1}(\boldsymbol{0}, 10\boldsymbol{\Sigma}_u)$ where $\boldsymbol{\Sigma}_u = (1-\rho_u)\boldsymbol{I}_{p-1} + \rho_u \boldsymbol{j}_{p-1}\boldsymbol{j}_{p-1}'$ for $\rho_u = 0.6$ and $\boldsymbol{j}_{p-1} = (1, \ldots, 1)' \in \boldsymbol{R}^{p-1}$. Let d_i's in the model (1.1) be generated as $d_i = 1/[1 + Bin(10, 1/2)]$ for $i = 1, \ldots, k$, where $Bin(10, 1/2)$ is a random variable distributed as a binomial distribution with mean 5 and success probability 1/2. The regression coefficients $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_{p-1})'$ are set up as $\beta_i = 5(-1)^i(U_i + 1)$ for $i = 0, \ldots, p-1$ where U_i is a random number from a uniform distribution on the interval $(0, 1)$. Then, the observation vector $\boldsymbol{y}$ is generated from $\mathcal{N}_k(\boldsymbol{X}\boldsymbol{\beta}, \psi\boldsymbol{I}_k + \boldsymbol{D})$.

In the simulation experiments, we handle the two cases: (A) $k = 30, 10$, $p = 3$, $q = 2$, $\alpha = 5\%$ and $H_0 : \beta_1 = \beta_2 = 0$, (B) $k = 20$, $p = 6$, $q = 4$, $\alpha = 5\%, 1\%$ and $H_0 : \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$, where ψ takes the values $\psi = 0, 0.2, 0.4, 0.6, 0.8, 1.0$. A set of observations of the regressor variables $\boldsymbol{X}$ is generated, and 10,000 observations of the response variable $\boldsymbol{y}$ are generated from the model (1.2). Sizes of test statistics can be approximated based on these simulation experiments.

We first handle the test statistic $T_{GLS}(0)$ with $\psi = 0$, the GLS test $T_{GLS}(\widehat{\psi}_{PR})$, the test $T_B(\widehat{\psi}_{PR})$ with the crude Bartlett correction (2.7), the Bartlett-type corrections $T_1^*(\widehat{\psi}_{PR})$ and $T_2^*(\widehat{\psi}_{PR})$ for the case (A). The sizes of these test statistics for the nominal significance level $\alpha = 5\%$ are reported in Table 1. From the table, it is seen that $T_{GLS}(0)$ is too bad, $T_{GLS}(\widehat{\psi}_{PR})$ is not good, and $T_B(\widehat{\psi}_{PR})$ is not good for $k = 10$, while the Bartlett-type correction GLS tests $T_1^*(\widehat{\psi}_{PR})$ and $T_2^*(\widehat{\psi}_{PR})$ are excellent. Especially for $k = 30$, their performances are very nice.

The sizes of the test statistics $T_{GLS}(\widehat{\psi})$, $T_1^*(\widehat{\psi})$ and $T_2^*(\widehat{\psi})$ for $\widehat{\psi} = \widehat{\psi}_{FH}$, $\widehat{\psi}_{ML}$ and $\widehat{\psi}_{REML}$ are reported in Table 2 in the case (A). Table 3 reports the sizes of the test statistics $T_{GLS}(\widehat{\psi})$, $T_1^*(\widehat{\psi})$ and $T_2^*(\widehat{\psi})$ for $\widehat{\psi} = \widehat{\psi}_{PR}$, $\widehat{\psi}_{FH}$, $\widehat{\psi}_{ML}$ and $\widehat{\psi}_{REML}$ in the case (B) for the nominal significance level $\alpha = 5\%, 1\%$. From Table 2, the Bartlett-type correction GLS tests $T_1^*(\widehat{\psi})$ and $T_2^*(\widehat{\psi})$ improve the GLS tests $T_{GLS}(\widehat{\psi})$ for $\widehat{\psi} = \widehat{\psi}_{FH}$, $\widehat{\psi}_{ML}$ and $\widehat{\psi}_{REML}$. The test $T_2(\widehat{\psi}_{ML})$ is too conservative in the case of $k = 10$. We can observe the same property of $T_2(\widehat{\psi}_{ML})$ in Table 3 for $k = 20$. From these tables, we can see that the Bartlett-type correction GLS tests $T_1^*(\widehat{\psi})$ and $T_2^*(\widehat{\psi})$ have very nice size properties for $k = 30$, while their size properties are still good for $k = 10, 20$ except $T_2^*(\widehat{\psi}_{ML})$. Tables 1-3 seem to suggest the use of $T_2^*(\widehat{\psi})$ for $\widehat{\psi} = \widehat{\psi}_{PR}$, $\widehat{\psi}_{FH}$ and $\widehat{\psi}_{REML}$ and the use of $T_1^*(\widehat{\psi})$ for $\widehat{\psi} = \widehat{\psi}_{ML}$.

4 Proof of Theorem 2.1

Recall that $\boldsymbol{W}(\psi)=\boldsymbol{C}'\left\{\boldsymbol{C}\{\boldsymbol{A}_1(\psi)\}^{-1}\boldsymbol{C}'\right\}^{-1}\boldsymbol{C}$, $\boldsymbol{W}^{(i)}(\psi)=(\mathrm{d}^i/\mathrm{d}\psi^i)\boldsymbol{W}(\psi)$ and $\boldsymbol{A}_i(\psi)=\boldsymbol{X}'\boldsymbol{\Sigma}^{-i}(\psi)\boldsymbol{X}$ for $i=1,2,3$. Let $\widehat{\boldsymbol{\beta}}^{(i)}(\psi)=\{\partial^i/\partial\psi^i\}\widehat{\boldsymbol{\beta}}(\psi)$ for i=1, 2. For notational convenience, we omit (ψ) in $\boldsymbol{W}(\psi)$, $\boldsymbol{W}^{(i)}(\psi)$, $\widehat{\boldsymbol{\beta}}(\psi)$, $\boldsymbol{A}_i(\psi)$, $\boldsymbol{\Sigma}(\psi)$ and others. Under the null hypothesis $H_0:\boldsymbol{C}\boldsymbol{\beta}=\boldsymbol{b}$, we want to derive the asymptotic expansion of the distribution of $T_{GLS}(\widehat{\psi})$. To this end, we first expand the characteristic function of $T_{GLS}(\widehat{\psi})$ given by $\phi(t)=E_{\boldsymbol{\beta}}[\exp\{itT_{GLS}(\widehat{\psi})\}]$, which is written as

$$\phi(t)=E_{\mathbf{0}}[\exp\{it\widehat{\boldsymbol{\beta}}(\widehat{\psi})'\boldsymbol{W}(\widehat{\psi})\widehat{\boldsymbol{\beta}}(\widehat{\psi})\}]=E_{\mathbf{0}}[\exp\{it\boldsymbol{y}'\boldsymbol{P}(\widehat{\psi})\boldsymbol{y}\}],$$

where $i=\sqrt{-1}$ and

$$\boldsymbol{P}(\psi)=\boldsymbol{P}=\boldsymbol{\Sigma}(\psi)^{-1}\boldsymbol{X}\boldsymbol{A}_1(\psi)^{-1}\boldsymbol{W}(\psi)\boldsymbol{A}_1(\psi)^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}(\psi)^{-1}. \tag{4.1}$$

Hereafter, we omit (ψ) in $\boldsymbol{W}(\psi)$, $\boldsymbol{P}(\psi)$, $\boldsymbol{\Sigma}(\psi)$, $\boldsymbol{A}_i(\psi)$ and others, and we can put $\boldsymbol{\beta}=\mathbf{0}$ without any loss of generality and omit $\mathbf{0}$ in the expectation notation $E_{\mathbf{0}}[\cdot]$.

Using Taylor series expansion of $\exp\{it\boldsymbol{y}'\boldsymbol{P}(\widehat{\psi})\boldsymbol{y}\}$ around $\widehat{\psi}=\psi$, we can approximate $\phi(t)$ as

$$\begin{aligned}\phi(t)=&E\Big[\exp\{it\boldsymbol{y}'\boldsymbol{P}\boldsymbol{y}\}\big[1+it\boldsymbol{y}'\boldsymbol{P}^{(1)}\boldsymbol{y}(\widehat{\psi}-\psi)\\&+2^{-1}\big\{it\boldsymbol{y}'\boldsymbol{P}^{(2)}\boldsymbol{y}+(it)^2(\boldsymbol{y}'\boldsymbol{P}^{(1)}\boldsymbol{y})^2\big\}(\widehat{\psi}-\psi)^2\big]\Big]+O(k^{-3/2}),\end{aligned} \tag{4.2}$$

For the sake of simplicity, we use the notation $E^*[\cdot]$ defined by

$$E^*[\boldsymbol{g}(\boldsymbol{y})]=E[\exp\{it\boldsymbol{y}'\boldsymbol{P}\boldsymbol{y}\}\boldsymbol{g}(\boldsymbol{y})]=C_k\int\boldsymbol{g}(\boldsymbol{y})\exp\{-\boldsymbol{y}'\boldsymbol{V}^{-1}\boldsymbol{y}\}\mathrm{d}\boldsymbol{y}, \tag{4.3}$$

for a function $\boldsymbol{g}(\boldsymbol{y})$, where $\boldsymbol{V}^{-1}=\boldsymbol{\Sigma}^{-1}-2it\boldsymbol{P}$ and $C_k=(2\pi)^{-k/2}|\boldsymbol{\Sigma}|^{-1/2}$. Then, we need to evaluate the terms $I_1=E^*[\boldsymbol{y}'\boldsymbol{P}^{(1)}\boldsymbol{y}(\widehat{\psi}-\psi)]$, $I_2=E^*[\boldsymbol{y}'\boldsymbol{P}^{(2)}\boldsymbol{y}(\widehat{\psi}-\psi)^2]$ and $I_3=E^*[(\boldsymbol{y}'\boldsymbol{P}^{(1)}\boldsymbol{y})^2(\widehat{\psi}-\psi)^2]$. The Stein identity proposed by Stein (1973, 81) is useful for the purpose. Since the Stein identity is based on the integration by parts, we can verify that the Stein identity under the notation $E^*[\cdot]$ is given by

$$E^*[\boldsymbol{y}'\boldsymbol{g}(\boldsymbol{y})]=E^*[\boldsymbol{\nabla}'[\boldsymbol{V}\boldsymbol{g}(\boldsymbol{y})]], \tag{4.4}$$

for k-dimensional absolutely continuous function $\boldsymbol{g}(\boldsymbol{y})$, where $\boldsymbol{\nabla}=\partial/\partial\boldsymbol{y}=(\partial/\partial y_1,\ldots,\partial/\partial y_k)'$ for $\boldsymbol{y}=(y_1,\ldots,y_k)'$. For scalar function $f(\boldsymbol{Y})$, the following equality is useful:

$$\boldsymbol{\nabla}'[\boldsymbol{g}(\boldsymbol{y})f(\boldsymbol{y})]=\{\boldsymbol{\nabla}'\boldsymbol{g}(\boldsymbol{y})\}f(\boldsymbol{y})+\boldsymbol{g}(\boldsymbol{y})'\boldsymbol{\nabla}f(\boldsymbol{y}). \tag{4.5}$$

Applying the Stein identity to I_1, we can rewrite it as

$$\begin{aligned}I_1=&E^*\Big[\boldsymbol{\nabla}'[\boldsymbol{V}\boldsymbol{P}^{(1)}\boldsymbol{y}(\widehat{\psi}-\psi)]\Big]\\=&E^*\Big[\mathrm{tr}\,[\boldsymbol{V}\boldsymbol{P}^{(1)}](\widehat{\psi}-\psi)\Big]+E^*\Big[\boldsymbol{y}'\boldsymbol{P}^{(1)}\boldsymbol{V}(\boldsymbol{\nabla}\widehat{\psi})\Big].\end{aligned} \tag{4.6}$$

The Stein identity is applied again to the second term to get that

$$E^*\Big[\boldsymbol{y}'\boldsymbol{P}^{(1)}\boldsymbol{V}(\boldsymbol{\nabla}\widehat{\psi})\Big] = E^*\Big[\boldsymbol{\nabla}'[\boldsymbol{V}\boldsymbol{P}^{(1)}\boldsymbol{V}(\boldsymbol{\nabla}\widehat{\psi})]\Big] = E^*\Big[\mathrm{tr}\,[\boldsymbol{V}\boldsymbol{P}^{(1)}\boldsymbol{V}(\boldsymbol{\nabla}\boldsymbol{\nabla}'\widehat{\psi})]\Big], \tag{4.7}$$

so that I_1 is expressed as

$$I_1 = \mathrm{tr}\,[\boldsymbol{V}\boldsymbol{P}^{(1)}]E^*[\widehat{\psi}-\psi] + E^*\Big[\mathrm{tr}\,[\boldsymbol{V}\boldsymbol{P}^{(1)}\boldsymbol{V}(\boldsymbol{\nabla}\boldsymbol{\nabla}'\widehat{\psi})]\Big]. \tag{4.8}$$

The same arguments as in (4.6) and (4.7) can be used to evaluate I_2 as

$$I_2 = \mathrm{tr}\,[\boldsymbol{V}\boldsymbol{P}^{(2)}]E^*[(\widehat{\psi}-\psi)^2] + 2E^*\Big[\mathrm{tr}\,[\boldsymbol{V}\boldsymbol{P}^{(2)}\boldsymbol{V}\boldsymbol{\Psi}]\Big], \tag{4.9}$$

where $\boldsymbol{\Psi} = (\boldsymbol{\nabla}\boldsymbol{\nabla}'\widehat{\psi})(\widehat{\psi}-\psi) + (\boldsymbol{\nabla}\widehat{\psi})(\boldsymbol{\nabla}'\widehat{\psi})$. Also, I_3 is rewritten as

$$\begin{aligned}I_3 =& \mathrm{tr}\,[\boldsymbol{V}\boldsymbol{P}^{(1)}]E^*[\boldsymbol{y}'\boldsymbol{P}^{(1)}\boldsymbol{y}(\widehat{\psi}-\psi)^2] + 2E^*\Big[\boldsymbol{y}'\boldsymbol{P}^{(1)}\boldsymbol{V}\boldsymbol{P}^{(1)}\boldsymbol{y}(\widehat{\psi}-\psi)^2\Big]\\ &+ 2E^*\Big[\boldsymbol{y}'\boldsymbol{P}^{(1)}\boldsymbol{V}(\boldsymbol{\nabla}\widehat{\psi})\boldsymbol{y}'\boldsymbol{P}^{(1)}\boldsymbol{y}(\widehat{\psi}-\psi)\Big].\end{aligned}$$

Using the same arguments as in (4.6) and (4.7) based on the Stein identity, we can evaluate I_3 as

$$\begin{aligned}I_3 =& (\mathrm{tr}\,[\boldsymbol{V}\boldsymbol{P}^{(1)}])^2E^*[(\widehat{\psi}-\psi)^2] + 2E^*[\mathrm{tr}\,[\boldsymbol{V}\boldsymbol{P}^{(1)}\boldsymbol{V}\boldsymbol{\Psi}]]\\ &+ 2\big\{\mathrm{tr}\,[\boldsymbol{V}\boldsymbol{P}^{(1)}\boldsymbol{V}\boldsymbol{P}^{(1)}]E^*[(\widehat{\psi}-\psi)^2] + 2E^*[\mathrm{tr}\,[\boldsymbol{V}\boldsymbol{P}^{(1)}\boldsymbol{V}\boldsymbol{P}^{(1)}\boldsymbol{V}\boldsymbol{\Psi}]\big\}\\ &+ 2\big\{\mathrm{tr}\,[\boldsymbol{V}\boldsymbol{P}^{(1)}]E^*[\mathrm{tr}\,[\boldsymbol{V}\boldsymbol{P}^{(1)}\boldsymbol{V}\boldsymbol{\Psi}]] + E^*[\mathrm{tr}\,[\boldsymbol{V}\boldsymbol{P}^{(1)}\boldsymbol{V}\boldsymbol{P}^{(1)}\boldsymbol{V}\boldsymbol{\Psi}]]\\ &\quad + E^*[\boldsymbol{y}'\boldsymbol{P}^{(1)}\boldsymbol{V}\boldsymbol{\Psi}\boldsymbol{V}\boldsymbol{P}^{(1)}\boldsymbol{y}]\big\}.\end{aligned} \tag{4.10}$$

Substituting $\boldsymbol{P}$ given in (4.1), we can simplify the terms I_1, I_2 and I_3 given in (4.8), (4.9) and (4.10). Note that

$$\boldsymbol{V} = \big(\boldsymbol{\Sigma}^{-1} - 2it\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\boldsymbol{A}_1^{-1}\boldsymbol{W}\boldsymbol{A}_1^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\big)^{-1} = \quad \boldsymbol{\Sigma} + (s-1)\boldsymbol{X}\boldsymbol{A}_1^{-1}\boldsymbol{W}\boldsymbol{A}_1^{-1}\boldsymbol{X}', \tag{4.11}$$

for $s = 1/(1-2it)$. Also note that $\boldsymbol{P}^{(1)}$ can be written as

$$\begin{aligned}\boldsymbol{P}^{(1)} =& (\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\boldsymbol{A}_1^{-1}\boldsymbol{A}_2 - \boldsymbol{\Sigma}^{-2}\boldsymbol{X})\boldsymbol{A}_1^{-1}\boldsymbol{W}\boldsymbol{A}_1^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\\ &+ \boldsymbol{\Sigma}^{-1}\boldsymbol{X}\boldsymbol{A}_1^{-1}\boldsymbol{W}\boldsymbol{A}_1^{-1}(\boldsymbol{A}_2\boldsymbol{A}_1^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1} - \boldsymbol{X}'\boldsymbol{\Sigma}^{-2})\\ &+ \boldsymbol{\Sigma}^{-1}\boldsymbol{X}\boldsymbol{A}_1^{-1}\boldsymbol{W}^{(1)}\boldsymbol{A}_1^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}.\end{aligned}$$

Since $\boldsymbol{W}\boldsymbol{A}_1^{-1}\boldsymbol{W} = \boldsymbol{W}$ and $\boldsymbol{W}\boldsymbol{A}_1^{-1}\boldsymbol{W}^{(1)} = \boldsymbol{W}^{(1)}$, it can be seen that

$$\begin{aligned}\boldsymbol{V}\boldsymbol{P}^{(1)} =& (\boldsymbol{X}\boldsymbol{A}_1^{-1}\boldsymbol{A}_2 - \boldsymbol{\Sigma}^{-1}\boldsymbol{X})\boldsymbol{A}_1^{-1}\boldsymbol{W}\boldsymbol{A}_1^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\\ &+ s\boldsymbol{X}\boldsymbol{A}_1^{-1}\boldsymbol{W}\boldsymbol{A}_1^{-1}(\boldsymbol{A}_2\boldsymbol{A}_1^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1} - \boldsymbol{X}'\boldsymbol{\Sigma}^{-2})\\ &+ s\boldsymbol{X}\boldsymbol{A}_1^{-1}\boldsymbol{W}^{(1)}\boldsymbol{A}_1^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1},\end{aligned} \tag{4.12}$$

which implies that

$$\begin{aligned}\mathrm{tr}\,[\boldsymbol{V}\boldsymbol{P}^{(1)}] =& s\mathrm{tr}\,[\boldsymbol{W}^{(1)}\boldsymbol{A}_1^{-1}],\\ \mathrm{tr}\,[(\boldsymbol{V}\boldsymbol{P}^{(1)})^2] =& s^2\mathrm{tr}\,[(\boldsymbol{W}^{(1)}\boldsymbol{A}_1^{-1})^2] + 2s\mathrm{tr}\,[(\boldsymbol{A}_3 - \boldsymbol{A}_2\boldsymbol{A}_1^{-1}\boldsymbol{A}_2)\boldsymbol{A}_1^{-1}\boldsymbol{W}\boldsymbol{A}_1^{-1}].\end{aligned} \tag{4.13}$$

From (4.13), it follows that $\partial\mathrm{tr}\,[\boldsymbol{V}\boldsymbol{P}^{(1)}]/\partial\psi = s\mathrm{tr}\,[\boldsymbol{W}^{(2)}\boldsymbol{A}_1^{-1}] + s\mathrm{tr}\,[\boldsymbol{W}^{(1)}\boldsymbol{A}_1^{-1}\boldsymbol{A}_2\boldsymbol{A}_1^{-1}]$. Since $\partial\mathrm{tr}\,[\boldsymbol{V}\boldsymbol{P}^{(1)}]/\partial\psi = \mathrm{tr}\,[\boldsymbol{V}^{(1)}\boldsymbol{P}^{(1)}] + \mathrm{tr}\,[\boldsymbol{V}\boldsymbol{P}^{(2)}]$ and

$$\begin{aligned}\boldsymbol{V}^{(1)} =& \boldsymbol{I}_k + (s-1)\boldsymbol{X}\boldsymbol{A}_1^{-1}\boldsymbol{W}^{(1)}\boldsymbol{A}_1^{-1}\boldsymbol{X}'\\ &+ (s-1)\boldsymbol{X}\boldsymbol{A}_1^{-1}\boldsymbol{A}_2\boldsymbol{A}_1^{-1}\boldsymbol{W}\boldsymbol{A}_1^{-1}\boldsymbol{X}' + (s-1)\boldsymbol{X}\boldsymbol{A}_1^{-1}\boldsymbol{W}\boldsymbol{A}_1^{-1}\boldsymbol{A}_2\boldsymbol{A}_1^{-1}\boldsymbol{X}',\end{aligned}$$

it can be observed that

$$\mathrm{tr}\,[\boldsymbol{V}\boldsymbol{P}^{(2)}] = s\mathrm{tr}\,[\boldsymbol{W}^{(2)}\boldsymbol{A}_1^{-1}] + 2\mathrm{tr}\,[(\boldsymbol{A}_3 - \boldsymbol{A}_2\boldsymbol{A}_1^{-1}\boldsymbol{A}_2)\boldsymbol{A}_1^{-1}\boldsymbol{W}\boldsymbol{A}_1^{-1}]. \tag{4.14}$$

It is also noted that

$$\begin{aligned}\boldsymbol{V}\boldsymbol{P}^{(1)}\boldsymbol{V} =& s(\boldsymbol{X}\boldsymbol{A}_1^{-1}\boldsymbol{A}_2 - \boldsymbol{\Sigma}^{-1}\boldsymbol{X})\boldsymbol{A}_1^{-1}\boldsymbol{W}\boldsymbol{A}_1^{-1}\boldsymbol{X}'\\ &+ s\boldsymbol{X}\boldsymbol{A}_1^{-1}\boldsymbol{W}\boldsymbol{A}_1^{-1}(\boldsymbol{A}_2\boldsymbol{A}_1^{-1}\boldsymbol{X}' - \boldsymbol{X}'\boldsymbol{\Sigma}^{-1})\\ &+ s^2\boldsymbol{X}\boldsymbol{A}_1^{-1}\boldsymbol{W}^{(1)}\boldsymbol{A}_1^{-1}\boldsymbol{X}',\end{aligned} \tag{4.15}$$

Then, from the condition (A2), it follows that $\mathrm{tr}\,[\boldsymbol{V}\boldsymbol{P}^{(1)}\boldsymbol{V}(\boldsymbol{\nabla}\boldsymbol{\nabla}'\widehat{\psi})] = 0$ and $\mathrm{tr}\,[\boldsymbol{V}\boldsymbol{P}^{(1)}\boldsymbol{V}\boldsymbol{\Psi}] = 0$. Similarly, from the condition (A3), we can verify that $\mathrm{tr}\,[\boldsymbol{V}\boldsymbol{P}^{(2)}\boldsymbol{V}\boldsymbol{\Psi}] = O_p(k^{-3/2})$, $\mathrm{tr}\,[\boldsymbol{V}\boldsymbol{P}^{(1)}\boldsymbol{V}\boldsymbol{P}^{(1)}\boldsymbol{V}\boldsymbol{\Psi}] = O_p(k^{-3/2})$ and $\boldsymbol{y}'\boldsymbol{P}^{(1)}\boldsymbol{V}\boldsymbol{\Psi}\boldsymbol{V}\boldsymbol{P}^{(1)}\boldsymbol{y} = O_p(k^{-3/2})$. Hence, we obtain the expressions that

$$\begin{aligned}I_1 =& s\,\mathrm{tr}\,[\boldsymbol{W}^{(1)}\boldsymbol{A}_1^{-1}]E^*[\widehat{\psi} - \psi],\\ I_2 =& \big\{s\,\mathrm{tr}\,[\boldsymbol{W}^{(2)}\boldsymbol{A}_1^{-1}] + 2\mathrm{tr}\,[(\boldsymbol{A}_3 - \boldsymbol{A}_2\boldsymbol{A}_1^{-1}\boldsymbol{A}_2)\boldsymbol{A}_1^{-1}\boldsymbol{W}\boldsymbol{A}_1^{-1}]\big\}E^*[(\widehat{\psi} - \psi)^2] + O(k^{-3/2}),\\ I_3 =& s^2\left\{\mathrm{tr}\,[\boldsymbol{W}^{(1)}\boldsymbol{A}_1^{-1}]\right\}^2E^*[(\widehat{\psi} - \psi)^2]\\ &+ 2\big\{s^2\,\mathrm{tr}\,[(\boldsymbol{W}^{(1)}\boldsymbol{A}_1^{-1})^2] + 2s\,\mathrm{tr}\,[(\boldsymbol{A}_3 - \boldsymbol{A}_2\boldsymbol{A}_1^{-1}\boldsymbol{A}_2)\boldsymbol{A}_1^{-1}\boldsymbol{W}\boldsymbol{A}_1^{-1}]\big\}E^*[(\widehat{\psi} - \psi)^2] + O(k^{-3/2}).\end{aligned}$$

From the result of Kackar and Harville (1984), it follows that $\widehat{\psi}$ is independent of $\widehat{\boldsymbol{\beta}}$ under the condition (A5). From the definition of $E^*[\cdot]$ and the fact that $\boldsymbol{y}'\boldsymbol{P}\boldsymbol{y} = \widehat{\boldsymbol{\beta}}'\boldsymbol{W}\widehat{\boldsymbol{\beta}}$, we can see that $E^*[\widehat{\psi} - \psi] = E[\exp\{it\widehat{\boldsymbol{\beta}}'\boldsymbol{W}\widehat{\boldsymbol{\beta}}\}(\widehat{\psi} - \psi)] = E[\exp\{it\widehat{\boldsymbol{\beta}}'\boldsymbol{W}\widehat{\boldsymbol{\beta}}\}]Bias(\widehat{\psi})$ and $E^*[(\widehat{\psi} - \psi)^2] = E[\exp\{it\widehat{\boldsymbol{\beta}}'\boldsymbol{W}\widehat{\boldsymbol{\beta}}\}]E[(\widehat{\psi} - \psi)^2]$. From (A4), note that $E[(\widehat{\psi} - \psi)^2] = Var(\widehat{\psi}) + o(k^{-1})$. Noting that $E[\exp\{it\widehat{\boldsymbol{\beta}}'\boldsymbol{W}\widehat{\boldsymbol{\beta}}\}] = s^{q/2}$ and substituting I_1, I_2 and I_3 into (4.3), we can get the approximation

$$\begin{aligned}\phi(t) =& E[\exp\{it\widehat{\boldsymbol{\beta}}'\boldsymbol{W}\widehat{\boldsymbol{\beta}}\}] + itI_1 + 2^{-1}itI_2 + 2^{-1}(it)^2I_3 + O(k^{-3/2})\\ =& s^{q/2} + \frac{s-1}{2}s^{q/2}\mathrm{tr}\,[\boldsymbol{W}^{(1)}\boldsymbol{A}_1^{-1}]Bias(\widehat{\psi})\\ &+ \frac{s-1}{2}s^{q/2}\left\{\frac{1}{2}\mathrm{tr}\,[\boldsymbol{W}^{(2)}\boldsymbol{A}_1^{-1}] + \mathrm{tr}\,[\boldsymbol{A}_1^{-1}\boldsymbol{W}\boldsymbol{A}_1^{-1}(\boldsymbol{A}_3 - \boldsymbol{A}_2\boldsymbol{A}_1^{-1}\boldsymbol{A}_2)\right\}Var(\widehat{\psi})\\ &+ \frac{(s-1)^2}{8}s^{q/2}\{(\mathrm{tr}\,[\boldsymbol{W}^{(1)}\boldsymbol{A}_1^{-1}])^2 + 2\mathrm{tr}\,[\boldsymbol{W}^{(1)}\boldsymbol{A}_1^{-1}]^2\}Var(\widehat{\psi}) + O(k^{-3/2}),\end{aligned}$$

which is expressed as

$$\phi(t) = s^{q/2} + s^{q/2}(s-1)h_1(\psi) + s^{q/2}(s-1)^2 h_2(\psi) + o(k^{-1}), \tag{4.16}$$

for the functions $h_1(\psi)$ and $h_2(\psi)$ given in (2.4). Inverting $\phi(t)$ yields the asymptotic expansion of the distribution function of $T_{GLS}(\widehat{\psi})$ given as

$$\begin{aligned} P[T_{GLS}(\widehat{\psi}) \le x] =& G_q(x) + \{G_{q+2}(x) - G_q(x)\}h_1(\psi) \\ &+ \{G_{q+4}(x) - 2G_{q+2} + G_q(x)\}h_2(\psi) + o(k^{-1}), \end{aligned} \tag{4.17}$$

where $G_q(x)$ is the distribution function of χ_q^2, namely, $G_q(x) = \int_0^x g_q(y)\mathrm{d}y$ for the pdf $g_q(y)$ of χ_q^2. Noting that $G_{q+2}(x) - G_q(x) = -2g_{q+2}(x) = -2(x/q)g_q(x)$, we can get the expression (2.3) from (4.17), and the proof of Theorem 2.1 is complete. ■

References

[1] Datta, G.S. and Lahiri, P. (2000). A unified measure of uncertainty of estimated best linear unbiased predictors in small area estimation problems. *Statist. Sinica*, **10**, 613-627.

[2] Datta, G.S., Rao, J.N.K. and Smith, D.D. (2005). On measuring the variability of small area estimators under a basic area level model. *Biometrika*, **92**, 183-196.

[3] Enoki, H. and Aoshima, M. (2006). Transformations with improved asymptotic approximations and their accuracy. *SUT Journal of Mathematics*, Tokyo University of Science, **42**, 97-122.

[4] Fay, R.E. and Herriot, R. (1979). Estimates of income for small places: An application of James-Stein procedures to census data. *J. Amer. Statist. Assoc.*, **74**, 269-277.

[5] Fujikoshi, Y. (2000). Transformations with improved chi-squared approximations. *J. Multivariate Analysis*, **72**, 249-263.

[6] Fujisawa, H. (1997). Improvement on chi-squared approximation. *J. Multivariate Analysis*, **60**, 84-89.

[7] Kackar, R.N., and Harville, D.A. (1984). Approximations for standard errors of estimators of fixed and random effects in mixed linear models. *J. American Statist. Assoc.*, **79**, 853-862.

[8] Prasad, N.G.N. and Rao, J.N.K. (1990). The estimation of the mean squared error of small-area estimators. *J. American Statist. Assoc.*, **85**, 163-171.

[9] Rao, J.N.K, Sutradhar, B.C., and Yue, K. (1993). Generalized least squares F test in regression analysis with two-stage cluster samples. *J. American Statist. Assoc.*, **88**, 1388-1391.

[10] Stein, C. (1973). Estimation of the mean of a multivariate normal distribution. In *Proc. Prague Symp. Asymptotic Statist.*, 345-381.

[11] Stein, C. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.*, **9**, 1135-1151.

[12] Wu, C.F.J, Holt, D., and Holmes, D.J. (1988). The effect of two-stage sampling on the F statistic. *J. American Statist. Assoc.*, **83**, 150-159.

Table 1: Size Estimates (%) of Tests $T_{GLS}(0)$, $T_{GLS}(\widehat{\psi})$, $T_B(\widehat{\psi})$, $T_1^*(\widehat{\psi})$ and $T_2^*(\widehat{\psi})$ for $\widehat{\psi} = \widehat{\psi}_{PR}$ where $p = 3$, $q = 2$, $\alpha = 5\%$ and $H_0 : \beta_1 = \beta_2 = 0$

ψ	$T_{GLS}(0)$	$T_{GLS}(\widehat{\psi}_{PR})$	$T_B(\widehat{\psi}_{PR})$	$T_1^*(\widehat{\psi}_{PR})$	$T_2^*(\widehat{\psi}_{PR})$
			$k = 30$		
0.0	5.0	3.9	2.7	2.2	2.0
0.2	29.7	6.9	5.8	5.1	4.9
0.4	46.9	6.8	5.8	5.1	4.9
0.6	57.8	6.7	5.8	5.1	4.9
0.8	65.1	6.6	5.8	5.1	4.9
1.0	70.5	6.6	5.8	5.1	4.9
			$k = 10$		
0.0	5.1	3.6	1.4	0.4	0.0
0.2	29.0	10.6	6.8	4.1	1.3
0.4	45.9	11.8	8.1	5.8	3.2
0.6	56.7	11.9	8.5	6.3	4.1
0.8	63.7	11.8	8.7	6.6	4.7
1.0	68.7	11.8	8.7	6.6	4.9

Table 2: Size Estimates (%) of Tests $T_{GLS}(\widehat{\psi})$, $T_1^*(\widehat{\psi})$ and $T_2^*(\widehat{\psi})$ for $\widehat{\psi} = \widehat{\psi}_{FH}$, $\widehat{\psi}_{ML}$ and $\widehat{\psi}_{REML}$ where $p = 3$, $q = 2$, $\alpha = 5\%$ and $H_0 : \beta_1 = \beta_2 = 0$

	$\widehat{\psi}_{FH}$			$\widehat{\psi}_{ML}$			$\widehat{\psi}_{REML}$		
ψ	T_{GLS}	T_1^*	T_2^*	T_{GLS}	T_1^*	T_2^*	T_{GLS}	T_1^*	T_2^*
				$k = 30$					
0.0	4.0	2.5	2.4	4.0	1.6	1.5	3.7	2.3	2.2
0.2	6.8	5.1	5.0	8.9	4.8	4.7	6.9	5.1	4.9
0.4	6.8	5.1	5.0	8.7	5.0	4.8	6.8	5.1	5.0
0.6	6.7	5.1	5.0	8.7	5.0	4.8	6.7	5.1	5.0
0.8	6.7	5.1	5.0	8.6	5.0	4.8	6.7	5.1	5.0
1.0	6.6	5.1	5.0	8.6	5.0	4.9	6.7	5.1	5.0
				$k = 10$					
0.0	3.6	0.8	0.2	2.5	0.0	0.0	1.5	0.2	0.1
0.2	10.4	4.9	2.8	14.5	1.0	0.0	7.8	3.5	1.9
0.4	11.7	6.4	4.3	18.4	2.5	0.0	10.2	5.3	3.5
0.6	11.8	6.6	4.9	19.7	3.5	0.0	11.1	6.1	4.4
0.8	11.7	6.6	5.1	20.0	4.0	0.1	11.3	6.4	4.8
1.0	11.7	6.6	5.2	20.1	4.4	0.2	11.4	6.5	5.0

Table 3: Size Estimates (%) of Tests $T_{GLS}(\widehat{\psi})$, $T_1^*(\widehat{\psi})$ and $T_2^*(\widehat{\psi})$ for $\widehat{\psi} = \widehat{\psi}_{PR}$, $\widehat{\psi}_{FH}$, $\widehat{\psi}_{ML}$ and $\widehat{\psi}_{REML}$ where $k = 20$, $p = 6$, $q = 4$ and $H_0 : \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$

	$\widehat{\psi}_{PR}$			$\widehat{\psi}_{FH}$			$\widehat{\psi}_{ML}$			$\widehat{\psi}_{REML}$		
ψ	T_{GLS}	T_1^*	T_2^*	T_{GLS}	T_1^*	T_2^*	T_{GLS}	T_1^*	T_2^*	T_{GLS}	T_1^*	T_2^*
$\alpha = 5\%$												
0.0	3.7	1.0	0.6	3.7	1.3	0.9	3.1	0.0	0.0	2.2	0.6	0.4
0.2	10.3	5.6	4.4	10.1	5.8	4.8	19.9	3.1	1.7	9.6	5.5	4.5
0.4	10.5	6.2	5.2	10.3	6.3	5.4	21.7	4.6	3.0	10.2	6.1	5.1
0.6	10.4	6.3	5.3	10.3	6.3	5.4	21.9	4.8	3.2	10.3	6.3	5.3
0.8	10.4	6.3	5.4	10.3	6.2	5.4	21.9	5.0	3.3	10.2	6.2	5.4
1.0	10.4	6.3	5.5	10.3	6.2	5.5	21.7	5.0	3.3	10.2	6.2	5.4
$\alpha = 1\%$												
0.0	0.7	0.0	0.0	0.6	0.0	0.0	0.5	0.0	0.0	0.3	0.0	0.0
0.2	4.0	1.0	0.2	3.9	1.2	0.6	9.3	0.3	0.0	3.6	1.0	0.5
0.4	4.3	1.5	0.6	4.1	1.6	0.9	11.0	0.8	0.0	4.0	1.5	0.8
0.6	4.2	1.6	0.9	4.1	1.7	1.0	11.1	0.9	0.0	4.1	1.6	0.9
0.8	4.2	1.6	0.9	4.1	1.7	1.0	11.0	1.0	0.0	4.1	1.5	1.0
1.0	4.1	1.7	1.0	4.1	1.7	1.0	11.0	1.0	0.0	4.1	1.6	1.0